A Bayesian approach to the linking of key comparisons

M Krystek

Physikalisch-Technische Bundesanstalt,
Bundesallee 100, D-38116 Braunschweig, Germany

E-mail: Michael.Krystek@ptb.de

H Bosse

Physikalisch-Technische Bundesanstalt,
Bundesallee 100, D-38116 Braunschweig, Germany

E-mail: Harald.Bosse@ptb.de

Abstract. This contribution presents a Bayesian approach to the issue of linking of the results from key comparison measurements. A mathematical treatment based on Bayesian statistics for the analysis of the results from two comparisons with some joint participants is described. This robust statistical analysis provides expressions and standard uncertainties for the key comparison reference value (KCRV) and the degree of equivalence (DOE) as well as a conformity check without any assumption on a priority of one of the comparisons. In addition to the derivation of the mathematical formulae to be used for this type of 'distributed linking', we also present one synthetic and one real linking example and discuss possible applications of this new linking procedure.

1. Introduction

The CIPM Mutual Recognition Arrangement (MRA) [1], which was put into force in October 1999 and meanwhile was signed by 95 National Metrology Institutes (NMIs) and 4 international organisations worldwide describes the foundations and requirements for the mutual recognition of national measurement standards and for calibration and measurement certificates issued by NMIs. In addition the CIPM MRA covers a further 150 institutes designated by the signatory bodies. The MRA, chapter 3 states, that "...*the technical basis of this arrangement is the set of results obtained in the course of time through key comparisons carried out by the Consultative Committees of the CIPM, the BIPM, and the regional metrology organisations (RMOs), and published by the BIPM and maintained in the key comparison database. Key comparisons carried out by Consultative Committees or the BIPM are referred to as CIPM key comparisons; key comparisons carried out by regional metrology organisations are referred to as RMO key comparisons; RMO key comparisons must be linked to the corresponding CIPM key comparisons by means of joint participants. The degree of equivalence derived from an RMO key comparison has the same status as that derived from a CIPM key comparison.*" To further support mutual confidence in the validity of calibration and measurement certificates issued by participating institutes so-called supplementary comparisons shall be carried out in addition. The metrology community has

followed these documented requirements to carry out international comparison measurements [2] on a regular basis, whose results are continuously registered and made available to the public in the BIPM key comparison database (KCDB) [3].

The issue of linking by means of joint participants is already addressed in the MRA, however without prescribing a particular mathematical linking procedure. Different approaches for robust statistical linking of the results of comparisons, e. g. by transferring a key comparison reference value (KCRV) from one comparison to the other, have been discussed and proposed in literature and have been applied for the analysis of comparison results, too [4], [6], [7], [8].

Most of the proposed approaches for robust statistical linking follow the basic assumption of the MRA, namely that the results of an RMO key comparison have to be linked to the results of a CIPM key comparison. The CIPM key comparison in this concept generally is regarded as a primary comparison, both with respect to the chronological order and with respect to the participating laboratories, because "...*participation in a CIPM key comparison is open to laboratories having the highest technical competence and experience, normally the member laboratories of the appropriate Consultative Committee.*" (see sect. 6.1 of [1]). In this sense the linking of CIPM key comparisons and RMO key comparisons can be regarded as a hierarchical process. However, in [4] approaches for linking of comparisons through global minimisation by application of generalised least squares methods [5] were shortly discussed too, which can be regarded as examples of non-hierarchical linking methods.

In the field of dimensional metrology which is under the responsibility of the CIPM Consultative Committee for Length (CCL), international comparisons were started immediately after the MRA was signed. The results of the first CIPM key comparison on gauge blocks e. g. were already published in 2002 [9].

Based on the experience gained from the international comparisons in the field of dimensional metrology, the CCL working group on dimensional metrology (CCL-WGDM‡) has published a document which describes the characteristics of dimensional comparisons and which also proposes an alternative comparison scheme, the so-called CCL-RMO key comparison scheme [10]. The background and the details of the proposed comparison scheme are not repeated here, however some core statements from this document are cited:

- The CCL-RMO key comparisons follow the idea of several comparisons mutually linked together, without the necessity of a top level comparison delivering a KCRV.
- The classical scheme of CCL and RMO key comparisons can be considered to be a special case of the more general CCL-RMO scheme.
- In terms of linking laboratories within and across regions by calculation of their respective degrees of equivalence, the classical (hierarchical) and the CCL-RMO comparison scheme can be regarded to be equivalent. The linking of the comparisons is guaranteed by common participation of selected laboratories.

In [11] a generalised formalism was presented for linking a CIPM key comparison with similar regional key comparisons. As an example, this procedure was applied to link the results of a regional (SIM) gauge block comparison [12] to the results of the first CIPM key comparison on gauge blocks [9]. In the applied procedure, a so-called linking invariant parameter is calculated based on the results of the linking laboratories, i. e. those laboratories which participated in both comparisons, which in this case used different transfer standards.

This proposed method was discussed and well accepted in the CCL-WGDM, however it was also criticised that it still required to choose one comparison as primary. This condition

---

‡ In June 2009 the working group structure of CCL has changed: MRA related work is now dealt with by the WG-MRA while linking issues are addressed by the task group on on KC linking (TG-L)

is not in full alignment of the linking procedure to be applied for the analysis of comparisons according to the proposed CCL-RMO key comparison scheme. It was argued, that an adapted 'distributed linking' procedure should be developed which should provide a statistically rigorous linking taking into account all the available information of the comparisons to be linked, however without the necessity to choose one comparison as a primary one.

In this paper a Bayesian approach for implementation of such a 'distributed linking' is presented and an example for linking of two gauge block comparisons is discussed, which refers to the data published in [9],[12].

2. The scenario

Let us assume the following scenario: A group of laboratories (group A) has participated in a key comparison of a travelling standard A and another group of laboratories (group B) has participated in a key comparison of a travelling standard B. The task of each group was to measure one single measurand of the respective travelling standard (for example the length of a gauge block). In order to link the results of the two key comparisons, a certain group of laboratories (group C) has participated in both key comparisons, i. e. the laboratories of this group have been members of group A as well as of group B and thus have measured both travelling standards. All laboratories have reported the respective measurement results and the standard uncertainties associated with them. In addition, the laboratories of group C have reported the covariances associated with the measurement results of the two different travelling standards. It should further be assumed, that the standards have been stable, i. e. that they have not changed their properties in terms of time during the key comparisons.

In order to simplify the organisation of the data, we assume a labelling of the laboratories by assigning, independently of the membership in one of the groups, a unique natural number to each of them, starting by the number one in ascending order. In the following we will use the labels as indices for the respective data. It should be understood, although we are using an ordered set of indices, it is totally arbitrary which index is assigned to which laboratory. Any permutation of the indices used would also serve.

After having assigned the indices, we are able to refer to the index sets of the laboratories, rather than to the groups. Those laboratories, which have e. g. only measured the travelling standard A are within the index set $\mathfrak{I}_A \setminus \mathfrak{I}_B$.

3. The information available

After having assigned labels (indices) to the laboratories, we are able to summarise the available data in a clear way just by using the same indices for the laboratories and the data. In the following, we will denote the measurement results concerning the travelling standard A obtained by the laboratories of group A by $x_{A,i}$ $(i \in \mathfrak{I}_A)$ and the standard uncertainties associated with this results by $u(x_{A,i})$ $(i \in \mathfrak{I}_A)$. Accordingly, we will denote the measurement results concerning the travelling standard B obtained by the laboratories of group B by $x_{B,i}$ $(i \in \mathfrak{I}_B)$ and the standard uncertainties associated with this results by $u(x_{B,i})$ $(i \in \mathfrak{I}_B)$. It then only remains to deal with the covariances, associated with the measurement results concerning both travelling standards, as reported by the laboratories of the group C. In the following we will denote these covariances by $u(x_{A,i}, x_{B,i})$ $\left(i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)\right)$. Equivalently we may use the correlation coefficients, defined by

$$r_{AB,i} = \frac{u(x_{A,i}, x_{B,i})}{u(x_{A,i}) u(x_{B,i})}, \qquad i \in (\mathfrak{I}_A \cap \mathfrak{I}_B), \tag{1}$$

instead of the covariances. We will frequently use the correlation coefficients as abbreviations in the subsequent calculations.

Under the prerequisite, that the two travelling standards under consideration have been stable during the key comparison, the assumption is justified, that all measurement results are the experimental realisation of only two different quantities, i.e. the measurands of the travelling standards A and B, denoted in the following by Y_A and Y_B, respectively. In addition we assume, that the measurement results have already been corrected for possible systematic deviations in a suitable way and the uncertainties associated with the corrections have been included in the combined standard uncertainties associated with the corrected results.

Note, that we use upper case letters to denote the measurands and *not* the respective measured value of a measurand, which we will denote by lower case letters throughout the text. We will however not differentiate between a measurand and its value (an ideal value to be estimated, sometimes called *true value*) and use the same letter for both, since the respective meaning follows from the context.

4. Establishing key comparison reference values

4.1. Preliminary remarks

Using the ordinary arithmetic mean of the measured values to calculate a key comparison reference value (KCRV) can not be justified, because in doing so none of the known uncertainty values are taken into account and thus we do not make use of all available information, as required by the *Guide to the Expression of Uncertainty in Measurement* (GUM) [13]. Therefore we shall apply the Bayesian theory of measurement uncertainty [14], which is entirely based on Bayesian statistics and the principle of maximum entropy (PME). The idea of this theory is first of all to establish a joint probability density function (pdf) of the measurands under consideration by taking into account the data obtained by measurement or otherwise given, together with their associated uncertainties, as well as physical relations and prior information about the measurands. This joint pdf expresses the state of incomplete knowledge about the measurands and is subsequently used to calculate the expectations of the measurands and its associated covariance matrix.

4.2. The joint probability density function

Since the measurements of all laboratories participating in the key comparison are assumed to be independent, we can assign probability density functions to the data of each laboratory individually. These pdfs belong to two different types only, because the data types of the groups A and B, excluding the data of group C, are similar and will thus lead to the same type of pdf. The data type of group C, however, is dissimilar from the data types of the groups A and B, excluding the data of group C, and hence we have to expect to get a different type of pdf for this group.

In order to assign pdfs§ to the data, we follow a suggestion of Jaynes [16], [17], [18], to use the principle of maximum entropy (PME) for this purpose. The PME requires to maximise the relative information entropy [19], [20] by a suitable pdf under the constraints imposed on the pdf by the normalisation condition and the data. In the process we make the usual

---

§ We will use here and throughout the document the abbreviation

$$p(X|Y) = p_{X|Y}(x|y)$$

for the conditional pdf, in order to simplify the notation of the formulae.

assumption, that the measured data and the associated variances and covariances, respectively, may be equated with the respective expectations.

The PME yields under the circumstances at hand

$$p\left(x_{A,i}, u(x_{A,i}) \mid Y_A\right) = \frac{1}{u(x_{A,i})\sqrt{2\pi}} \exp\left(-\frac{(Y_A - x_{A,i})^2}{2u^2(x_{A,i})}\right), \qquad i \in \mathfrak{I}_A \setminus \mathfrak{I}_B, \qquad (2)$$

$$p\left(x_{B,i}, u(x_{B,i}) \mid Y_B\right) = \frac{1}{u(x_{B,i})\sqrt{2\pi}} \exp\left(-\frac{(Y_B - x_{B,i})^2}{2u^2(x_{B,i})}\right), \qquad i \in \mathfrak{I}_B \setminus \mathfrak{I}_A, \qquad (3)$$

and

$$p\left(x_{A,i}, x_{B,i}, u(x_{A,i}), u(x_{B,i}), r_{AB,i} \mid Y_A, Y_B\right) = \frac{1}{2\pi\, u(x_{A,i})u(x_{B,i})\sqrt{1 - r_{AB,i}^2}}$$

$$\times \exp\left(-\frac{1}{2(1 - r_{AB,i}^2)}\left[\frac{(Y_A - x_{A,i})^2}{u^2(x_{A,i})} - 2\frac{r_{AB,i}(Y_A - x_{A,i})(Y_B - x_{B,i})}{u(x_{A,i})u(x_{B,i})} + \frac{(Y_B - x_{B,i})^2}{u^2(x_{B,i})}\right]\right),$$

$$i \in \mathfrak{I}_A \cap \mathfrak{I}_B. \quad (4)$$

These pdfs can be esteemed as sampling distributions an thus be used to establish the likelihood function in the usual way. Note that the values $x_{A,i}$ ($i \in \mathfrak{I}_A$) and $x_{B,i}$ ($i \in \mathfrak{I}_B$) as well as their associated uncertainties $u(x_{A,i})$ and $u(x_{B,i})$, respectively, are given data, while Y_A and Y_B are unknown quantities.

The prior pdf $p(Y_A, Y_B)$ would allow us to express our knowledge about the quantities Y_A and Y_B. However, in order to be conservative, we use Jeffreys' prior [15] here, i. e. we assign a constant to this prior pdf. If more knowledge about Y_A and Y_B would be available, the prior could be changed accordingly.

Using Bayes' theorem we obtain the posterior pdf

$$p(Y_A, Y_B \mid \mathfrak{D}) = Ce^{-\chi^2/2}, \qquad (5)$$

with the set $\mathfrak{D}$ representing all known data,

$$\chi^2 = \sum_{i \in (\mathfrak{I}_A \setminus \mathfrak{I}_B)} \frac{(Y_A - x_{A,i})^2}{u^2(x_{A,i})} + \sum_{i \in (\mathfrak{I}_B \setminus \mathfrak{I}_A)} \frac{(Y_B - x_{B,i})^2}{u^2(x_{B,i})}$$

$$+ \sum_{i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)} \frac{1}{1 - r_{AB,i}^2}\left[\frac{(Y_A - x_{A,i})^2}{u^2(x_{A,i})} - 2\frac{r_{AB,i}(Y_A - x_{A,i})(Y_B - x_{B,i})}{u(x_{A,i})u(x_{B,i})} + \frac{(Y_B - x_{B,i})^2}{u^2(x_{B,i})}\right] \quad (6)$$

and a normalisation constant C. This is the posterior pdf of the quantities Y_A and Y_B given the data $\mathfrak{D}$.

After some algebraic transformations, equation (6) can be written as

$$\chi^2 = \frac{1}{1 - \tilde{r}_{AB}^2}\left(\frac{(Y_A - \hat{y}_A)^2}{u^2(\hat{y}_A)} - 2\tilde{r}_{AB}\frac{(Y_A - \hat{y}_A)(Y_B - \hat{y}_B)}{u(\hat{y}_A)u(\hat{y}_B)} + \frac{(Y_B - \hat{y}_B)^2}{u^2(\hat{y}_B)}\right) + q^2 \qquad (7)$$

with

$$\hat{y}_A = \frac{bs_1 + cs_2}{ab - c^2} \qquad (8)$$

$$\hat{y}_B = \frac{cs_1 + as_2}{ab - c^2} \qquad (9)$$

$$u(\hat{y}_A) = \sqrt{\frac{b}{ab - c^2}} \qquad (10)$$

$$u(\hat{y}_B) = \sqrt{\frac{a}{ab - c^2}} \qquad (11)$$

$$\tilde{r}_{AB} = \frac{c}{\sqrt{ab}} \qquad (12)$$

$$q^2 = \sum_{i \in (\mathfrak{I}_A \setminus \mathfrak{I}_B)} \frac{(x_{A,i} - \hat{y}_A)^2}{u^2(x_{A,i})} + \sum_{i \in (\mathfrak{I}_B \setminus \mathfrak{I}_A)} \frac{(x_{B,i} - \hat{y}_B)^2}{u^2(x_{B,i})}$$
$$+ \sum_{i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)} \frac{1}{1 - r_{AB,i}^2} \left(\frac{(x_{A,i} - \hat{y}_A)^2}{u^2(x_{A,i})} - 2\frac{r_{AB,i}(x_{A,i} - \hat{y}_A)(x_{B,i} - \hat{y}_B)}{u(x_{A,i})u(x_{B,i})} + \frac{(x_{B,i} - \hat{y}_B)^2}{u^2(x_{B,i})} \right) \qquad (13)$$

and the abbreviations

$$a = \sum_{i \in (\mathfrak{I}_A \setminus \mathfrak{I}_B)} \frac{1}{u^2(x_{A,i})} + \sum_{i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)} \frac{1}{(1 - r_{AB,i}^2)u^2(x_{A,i})} \qquad (14)$$

$$b = \sum_{i \in (\mathfrak{I}_B \setminus \mathfrak{I}_A)} \frac{1}{u^2(x_{B,i})} + \sum_{i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)} \frac{1}{(1 - r_{AB,i}^2)u^2(x_{B,i})} \qquad (15)$$

$$c = \sum_{i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)} \frac{r_{AB,i}}{(1 - r_{AB,i}^2)u(x_{A,i})u(x_{B,i})} \qquad (16)$$

$$s_1 = \sum_{i \in (\mathfrak{I}_A \setminus \mathfrak{I}_B)} \frac{x_{A,i}}{u^2(x_{A,i})} + \sum_{i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)} \frac{1}{1 - r_{AB,i}^2} \left(\frac{x_{A,i}}{u^2(x_{A,i})} - \frac{r_{AB,i}x_{B,i}}{u(x_{A,i})u(x_{B,i})} \right) \qquad (17)$$

$$s_2 = \sum_{i \in (\mathfrak{I}_B \setminus \mathfrak{I}_A)} \frac{x_{B,i}}{u^2(x_{B,i})} + \sum_{i \in (\mathfrak{I}_A \cap \mathfrak{I}_B)} \frac{1}{1 - r_{AB,i}^2} \left(\frac{x_{B,i}}{u^2(x_{B,i})} - \frac{r_{AB,i}x_{A,i}}{u(x_{A,i})u(x_{B,i})} \right) \qquad (18)$$

Note, that the auxiliary quantities a, b, c, s_1, s_2, and thereby the quantities $\hat{y}_A$, $\hat{y}_B$, $u(\hat{y}_A)$, $u(\hat{y}_B)$, $\tilde{r}_{AB}$ as well, depend only on the data obtained by the measurement. Thus q^2 does not depend on the quantities Y_A and Y_B and can hence be absorbed by the normalisation constant. Therefore, after renormalisation, we finally get

$$p(Y_A, Y_B | \mathfrak{D}) = \frac{1}{2\pi u(\hat{y}_A)u(\hat{y}_B)\sqrt{1 - \tilde{r}_{AB}^2}}$$
$$\times \exp\left[-\frac{1}{2(1 - \tilde{r}_{AB}^2)} \left(\frac{(Y_A - \hat{y}_A)^2}{u^2(\hat{y}_A)} - 2\tilde{r}_{AB}\frac{(Y_A - \hat{y}_A)(Y_B - \hat{y}_B)}{u(\hat{y}_A)u(\hat{y}_B)} + \frac{(Y_B - \hat{y}_B)^2}{u^2(\hat{y}_B)} \right) \right], \qquad (19)$$

i. e. the posterior pdf of the quantities Y_A and Y_B is a bivariate Gaussian pdf.

4.3. The key comparison reference values

Since the posterior pdf of the quantities Y_A and Y_B is a bivariate Gaussian pdf, it follows that

$$\mathsf{E}[Y_A] = \hat{y}_A \qquad\qquad \mathsf{E}[Y_B] = \hat{y}_B \qquad (20)$$

are the expectations of the quantities Y_A and Y_B, respectively, and

$$\mathbf{U}_Y = \begin{pmatrix} u^2(\hat{y}_A) & \tilde{r}_{AB} u(\hat{y}_A) u(\hat{y}_B) \\ \tilde{r}_{AB} u(\hat{y}_A) u(\hat{y}_B) & u^2(\hat{y}_B) \end{pmatrix} \tag{21}$$

their associated covariance matrix. We now identify the values $\hat{y}_A$ and $\hat{y}_B$ as the key comparison reference values and the matrix $\mathbf{U}_Y$ as their associated covariance matrix.

For convenience we summarise the results, using the covariances rather than the correlation coefficients.

$$a = \sum_{i \in (\mathfrak{J}_A \setminus \mathfrak{J}_B)} \frac{1}{u^2(x_{A,i})} + \sum_{i \in (\mathfrak{J}_A \cap \mathfrak{J}_B)} \frac{u^2(x_{B,i})}{u^2(x_{A,i}) u^2(x_{B,i}) - u^2(x_{A,i}, x_{B,i})}, \tag{22}$$

$$b = \sum_{i \in (\mathfrak{J}_B \setminus \mathfrak{J}_A)} \frac{1}{u^2(x_{B,i})} + \sum_{i \in (\mathfrak{J}_A \cap \mathfrak{J}_B)} \frac{u^2(x_{A,i})}{u^2(x_{A,i}) u^2(x_{B,i}) - u^2(x_{A,i}, x_{B,i})}, \tag{23}$$

$$c = \sum_{i \in (\mathfrak{J}_A \cap \mathfrak{J}_B)} \frac{u(x_{A,i}, x_{B,i})}{u^2(x_{A,i}) u^2(x_{B,i}) - u^2(x_{A,i}, x_{B,i})}, \tag{24}$$

$$s_1 = \sum_{i \in (\mathfrak{J}_A \setminus \mathfrak{J}_B)} \frac{x_{A,i}}{u^2(x_{A,i})} + \sum_{i \in (\mathfrak{J}_A \cap \mathfrak{J}_B)} \frac{u^2(x_{B,i}) x_{A,i} - u(x_{A,i}, x_{B,i}) x_{B,i}}{u^2(x_{A,i}) u^2(x_{B,i}) - u^2(x_{A,i}, x_{B,i})}, \tag{25}$$

$$s_2 = \sum_{i \in (\mathfrak{J}_B \setminus \mathfrak{J}_A)} \frac{x_{B,i}}{u^2(x_{B,i})} + \sum_{i \in (\mathfrak{J}_A \cap \mathfrak{J}_B)} \frac{u^2(x_{A,i}) x_{B,i} - u(x_{A,i}, x_{B,i}) x_{A,i}}{u^2(x_{A,i}) u^2(x_{B,i}) - u^2(x_{A,i}, x_{B,i})}. \tag{26}$$

Key comparison reference values (KCRVs):

$$\hat{y}_A = \frac{b s_1 + c s_2}{ab - c^2}, \tag{27}$$

$$\hat{y}_B = \frac{c s_1 + a s_2}{ab - c^2}. \tag{28}$$

Standard uncertainties of the KCRVs:

$$u(\hat{y}_A) = \sqrt{\frac{b}{ab - c^2}}, \tag{29}$$

$$u(\hat{y}_B) = \sqrt{\frac{a}{ab - c^2}}. \tag{30}$$

Covariance of the KCRVs:

$$u(\hat{y}_A, \hat{y}_B) = \frac{c}{ab - c^2}. \tag{31}$$

5. Degrees of equivalence

The degrees of equivalence (DOE) are defined to be

$$d_{A,i} = x_{A,i} - \hat{y}_A, \qquad i \in \mathfrak{J}_A, \tag{32}$$

and

$$d_{B,i} = x_{B,i} - \hat{y}_B, \qquad i \in \mathfrak{J}_B. \tag{33}$$

These values give the deviations of each laboratory from the respective KCRV.

Following the rules for the propagation of variances, we obtain

$$u(d_{A,i}) = \sqrt{u^2(x_{A,i}) - u^2(\hat{y}_A)}, \qquad i \in \mathfrak{I}_A, \tag{34}$$

and

$$u(d_{B,i}) = \sqrt{u^2(x_{B,i}) - u^2(\hat{y}_B)}, \qquad i \in \mathfrak{I}_B, \tag{35}$$

for the respective uncertainties. The negative sign under the square root is due to the correlations $u(\hat{y}_A, x_{B,k}) = u^2(\hat{y}_A)$ and $u(\hat{y}_B, x_{B,k}) = u^2(\hat{y}_B)$ as shown in the appendix.

6. Conformity test

After the evaluation procedure has been accomplished, we have to check, whether the determined results do conform to the data given. This can be achieved in the following way. We observe, that on the one hand χ^2 is represented by equation (6) and on the other hand by equation (7). Thus, we can calculate the expectation $\mathsf{E}\left[\chi^2\right]$ in two different ways, yielding

$$\mathsf{E}\left[\chi^2\right] = N = 2 + q^2, \tag{36}$$

with

$$N = \mathrm{card}\,\mathfrak{I}_A + \mathrm{card}\,\mathfrak{I}_B \tag{37}$$

and q^2 given by equation (13). Thus, we obtain

$$q^2 = N - 2. \tag{38}$$

Note that the right hand side of this equation may be regarded as the number of the degrees of freedom, because it is just the total number of measured data diminished by the number of estimated quantities.

The condition (38) can, however, not be expected to be strictly fulfilled, but $N - 2$ is the most probable value of q^2, because if the data are actually statistically distributed according to the model and the measurement uncertainties given, equation (38) is the consequence. A significant deviation from this result is a signal that either the model is wrong or the data are suspect. Assuming the model to be correct, an outcome $q^2 > N - 2$ leads to the conclusion, that there is a strong possibility that either some of the measurement uncertainties have been underestimated or the respective measurement results contain uncorrected (or unknown) systematic deviations. Thus

$$q^2 \leq N - 2 \tag{39}$$

can be taken as an indication of the conformity of the estimated results with the data.

7. Examples

In order to demonstrate the application of the method proposed in this paper, we show two examples. Our first example is a synthetic one. Data have been produced by a simulated sampling from Gaussian distributions as given in equations (2) to (4), using the parameters‖ $Y_A = 110$, $\sigma_A = 20$, $Y_B = 120$, $\sigma_B = 50$, $\rho = 0,5$ and a sample size of $n = 50$. Subsequently the sampled data have been evaluated by using the usual formulae for the sample mean, the sample variance and the sample covariance. The resulting data thus obtained as well as their evaluation are shown in table 1 and table 2.

Table 1. Data of the synthetic example obtained from the sampled data by application of the usual formulae for the sample mean, the sample variance and the sample covariance.

Institute	$x_{A,i}$	$u(x_{A,i})$	$x_{B,i}$	$u(x_{B,i})$	$r_{AB,i}$
LAB-01	113,4	2,9			
LAB-02	112,1	2,8			
LAB-03	113,0	2,5			
LAB-04	110,6	2,6			
LAB-05	109,4	2,4			
LAB-06	107,0	2,6			
LAB-07	104,7	2,8			
LAB-08	109,0	2,6			
LAB-09	111,0	2,4	120,1	6,5	0,8
LAB-10	109,4	2,8	117,3	7,3	0,8
LAB-11	111,1	2,8	125,0	6,4	0,8
LAB-12	115,3	2,4	135,7	6,7	0,7
LAB-13			129,7	6,1	
LAB-14			129,1	7,5	
LAB-15			125,0	7,1	
LAB-16			123,6	6,6	
LAB-17			123,0	6,9	

As can be seen, the data do pass the conformity test. However, the KCRVs are slightly larger than the values used for the simulation. This is caused by the fact, that the correlation coefficients estimated from the sampled data tend to be too large compared to the correlation coefficient used for the simulation. This demonstrates the effect of small sample sizes.

As a second example data from the CCL-K1 [9] and SIM.L-K1 [12] gauge block comparisons have been evaluated. The three institutes CENAM, NIST and NRC participated in both comparisons and hence serve as linking laboratories. For convenience, the results and their associated standard uncertainties, as reported in the respective publications, are repeated in table 3. Values for the covariances have unfortunately not been reported, although systematic deviations present when measuring the gauge blocks in each of the linking laboratories with the same measuring system inevitably cause correlations, even if a correction for these systematic effects, as usually good practise, has been applied. Thus, following the recommendation given in the GUM in lack of this information¶, all covariances have been assumed to be zero, knowing that this is certainly not true.

The degrees of equivalence for the deviation from the nominal values and their associated standard uncertainties, as well as the key comparison reference values and their associated standard uncertainties for the groups A and B, respectively, obtained by the application of the linking method proposed in this paper are given in table 4 for the original data. It emerges, however, that the data do not pass the conformity test.

Assuming that the evaluation model is correct and that the respective measurement results have been corrected for systematic deviations, the test result leads to the conclusion, that there is a strong possibility that some of the measurement uncertainties have been

‖ σ and ρ denote the standard deviation and the correlation coefficient, respectively, of a Gaussian distribution.
¶ It should be noted, that this recommendation is consistent with the *modus operandi* of the Bayesian theory, where only *known* information is taken into account.

Table 2. Degrees of equivalence for the deviation from the nominal values and their associated standard uncertainties, as well as the key comparison reference values and their associated standard uncertainties for the groups A and B, respectively, obtained by the application of the linking method proposed in this paper for the synthetic example.

Institute	$d_{A,i}$	$u(d_{A,i})$	$d_{B,i}$	$u(d_{B,i})$
LAB-01	2,491	2,815		
LAB-02	1,191	2,712		
LAB-03	2,091	2,401		
LAB-04	-0,309	2,505		
LAB-05	-1,509	2,296		
LAB-06	-3,909	2,505		
LAB-07	-6,209	2,712		
LAB-08	-1,909	2,505		
LAB-09	0,091	2,296	-3,779	6,196
LAB-10	-1,509	2,712	-6,579	7,030
LAB-11	0,191	2,712	1,121	6,091
LAB-12	4,391	2,296	11,821	6,405
LAB-13			5,821	5,775
LAB-14			5,221	7,238
LAB-15			1,121	6,822
LAB-16			-0,279	6,300
LAB-17			-0,879	6,614

$\hat{y}_A = 110,909 \,, u(\hat{y}_A) = 0,698$

$\hat{y}_B = 123,879 \,, u(\hat{y}_B) = 1,966$

$q^2/(N-2) = 0,89$

underestimated. Inspecting the uncertainty values given in table 3, it is obvious that "...*INMETRO1 submitted very optimistic uncertainty claims relative to conventional capabilities*" [12] (the corresponding value is marked in bold face in the table). Since the gauge blocks have been measured at INMETRO by two completely different instruments, involving different staff members (data denoted by INMETRO1 represent results from a research grade instrument, while those denoted by INMETRO2 represent results from the routine gauge block calibration service offered by INMETRO) [12], it is reasonable to question the very low uncertainty stated for INMETRO1. Therefore, this uncertainty value has been increased here from 4,0 nm to 11,2 nm, which is the smallest value in order that the data pass the conformity test. The respective results obtained with this change are given in table 5.

A comparison of the results given in table 4 with those given in table 5 shows, that the reference value for group B and its associated standard uncertainty has noticeably changed. This strongly emphasises, that it is imperative to do a *rigorous* conformity test. The E_n criterion, for example, did not reveal any problem with the data from INMETRO1 during the SIM.l-K1 comparison. In fact the values $|E_{95}|$ as well as $|E|$ for the evaluation of the 100 mm steel gauge block data have even been comparable to those for the NRC [12]. The usual χ^2 criterion at a 5 % confidence level also fails to indicate a problem with the data.

Table 3. Results for the deviations from the nominal length and their associated standard uncertainties as taken from the CCL-K1 [9] and SIM.L-K1 [12] gauge block comparison reports (steel gauge block, nominal length 100 mm).

Institute	$x_{A,i}$/nm	$u(x_{A,i})$/nm	$x_{B,i}$/nm	$u(x_{B,i})$/nm
METAS	-96,0	13,0		
NPL	-140,0	33,0		
BNM-LNE	-110,0	16,0		
KRISS	-104,3	20,6		
NRLM	-89,4	16,3		
VNIIM	-104,0	15,0		
CSIRO	-114,0	16,0		
NIM	-90,0	10,3		
NIST	-117,0	17,9	-100,0	18,0
CENAM	-119,0	18,7	-93,0	23,0
NRC	-126,0	24,0	-124,0	26,0
INMETRO1			-98,0	**4,0**
INMETRO2			-68,0	29,0
INTI			-104,0	21,0
CEM			-148,0	17,0

Table 4. Degrees of equivalence for the deviation from the nominal values and their associated standard uncertainties, as well as the key comparison reference values and their associated standard uncertainties for the groups A and B, respectively, obtained by the application of the linking method proposed in this paper (results obtained for the original data in table 3).

Institute	$d_{A,i}$/nm	$u(d_{A,i})$/nm	$d_{B,i}$/nm	$u(d_{B,i})$/nm
METAS	7,6	12,1		
NPL	-36,4	32,6		
BNM-LNE	-6,4	15,2		
KRISS	-0,7	20,0		
NRLM	14,2	15,6		
VNIIM	-0,4	14,2		
CSIRO	-10,4	15,2		
NIM	13,6	9,1		
NIST	-13,4	17,2	0,5	17,6
CENAM	-15,4	18,1	7,5	22,7
NRC	-22,4	23,5	-23,5	25,7
INMETRO1			2,5	1,7
INMETRO2			32,5	28,8
INTI			-3,5	20,7
CEM			-47,5	16,6
	$\hat{y}_A = -103,6$, $u(\hat{y}_A) = 4,9$ nm			
	$\hat{y}_B = -100,5$, $u(\hat{y}_B) = 3,6$ nm			
	$q^2/(N-2) = 1,07$ (conformity test failed)			

Table 5. Degrees of equivalence for the deviation from the nominal values and their associated standard uncertainties, as well as the key comparison reference values and their associated standard uncertainties for the groups A and B, respectively, obtained by the application of the linking method proposed in this paper (results obtained after the uncertainty of the institute INMETRO1 has been increased from 4,0 nm to 11,2 nm).

Institute	$d_{A,i}/\text{nm}$	$u(d_{A,i})/\text{nm}$	$d_{B,i}/\text{nm}$	$u(d_{B,i})/\text{nm}$
METAS	7,6	12,1		
NPL	-36,4	32,6		
BNM-LNE	-6,4	15,2		
KRISS	-0,7	20,0		
NRLM	14,2	15,6		
VNIIM	-0,4	14,2		
CSIRO	-10,4	15,2		
NIM	13,6	9,1		
NIST	-13,4	17,2	6,7	16,6
CENAM	-15,4	18,1	13,7	22,0
NRC	-22,4	23,5	-17,3	25,1
INMETRO1			8,7	8,9
INMETRO2			38,7	28,2
INTI			2,7	19,9
CEM			-41,3	15,6
	$\hat{y}_A = -103,6, u(\hat{y}_A) = 4,9$ nm			
	$\hat{y}_B = -106,7, u(\hat{y}_B) = 6,8$ nm			
	$q^2/(N-2) = 1,00$			

8. Conclusion

In this contribution we have derived formulae for the linking parameters key comparison reference value (KCRV) and degree of equivalence (DOE) for analysis of two comparisons based on Bayesian statistics, i.e. taking all known information into account but without the necessity to choose one comparison as a primary one. The applicability of the linking approach was demonstrated on an example from gauge block metrology.

The proposed approach can be described as a type of 'distributed linking', which has been looked for especially in the area of dimensional metrology within the working groups of the CCL.

This 'distributed linking' method can, for example, be used to compare the results of two comparisons which were started in parallel using different transfer standards and where it does not seem meaningful to choose one comparison to be primary (defining a KCRV) and the other secondary (linking its comparison results to the KCRV of the first). Such comparisons with more than one loop have been organised if the number of interested participants became rather large, like e. g. in the finished line scale comparison EUROMET.L-K7 with 31 participants [21].

In a strict Bayesian sense, using the 'distributed linking' approach takes all known information from both comparisons into account to determine the linking parameters, i. e. all the measurement results and their uncertainties. Instead, in application of the hierarchical linking approaches one could in principle argue that not all of the information is taken into account, because e. g. the results gained from a RMO key comparison are not taken into

account to reflect or re-analyse the results of a CIPM key comparison, although by the twofold involvement of the linking laboratories the level of knowledge is increased for both comparisons. On the other hand the results of a subsequent RMO comparison linked to a key comparison by the described 'distributed linking' approach would also change the KCRV of the (already finished) key comparison. We see that this would impose new challenges for the operation of the KCDB.

The described approach can be extended to link more than two comparisons. Moreover, additional knowledge, as e. g. results from prior measurements, may be taken into account by choosing a suitable prior pdf.

9. Acknowledgement

Appendix 1 (Calculation of covariances)

In equations (34) and (35) the relations $u(\hat{y}_A, x_{B,k}) = u^2(\hat{y}_A)$ and $u(\hat{y}_B, x_{B,k}) = u^2(\hat{y}_B)$, respectively, have implicitly been used. In order to show that they are valid, we introduce the estimators

$$\hat{Y}_A = \frac{bS_1 + cS_2}{ab - c^2} \tag{40}$$

and

$$\hat{Y}_B = \frac{cS_1 + aS_2}{ab - c^2}, \tag{41}$$

where

$$S_1 = \sum_{i \in (\Im_A \setminus \Im_B)} \frac{X_{A,i}}{u^2(x_{A,i})} + \sum_{i \in (\Im_A \cap \Im_B)} \frac{u^2(x_{B,i})X_{A,i} - u(x_{A,i}, x_{B,i})X_{B,i}}{u^2(x_{A,i})u^2(x_{B,i}) - u^2(x_{A,i}, x_{B,i})} \tag{42}$$

and

$$S_2 = \sum_{i \in (\Im_B \setminus \Im_A)} \frac{X_{B,i}}{u^2(x_{B,i})} + \sum_{i \in (\Im_A \cap \Im_B)} \frac{u^2(x_{A,i})X_{B,i} - u(x_{A,i}, x_{B,i})X_{A,i}}{u^2(x_{A,i})u^2(x_{B,i}) - u^2(x_{A,i}, X_{B,i})}. \tag{43}$$

are functions of the random variables $X_{A,i}$ $(i \in \Im_A)$ and $X_{B,i}$ $(i \in \Im_B)$. We assume the measured values $x_{A,i}$ $(i \in \Im_A)$ and $x_{B,i}$ $(i \in \Im_B)$ to be a realisation of these random variables.

Since $E[X_{A,i}] = Y_A$, $E[X_{B,i}] = Y_B$, $Var[X_{A,i}] = u^2(x_{A,i})$, $Var[X_{B,i}] = u^2(x_{B,i})$ and $Cov[X_{A,i}, X_{B,i}] = u(x_{A,i}, x_{B,i})$ $(i \in \Im_A, i \in \Im_B)$ it can be verified, that

$$E[\hat{S}_1] = aY_A - cY_B, \qquad E[\hat{S}_2] = bY_B - cY_A, \tag{44}$$

as well as

$$Var[\hat{S}_1] = a, \qquad Var[\hat{S}_2] = b, \qquad Cov[\hat{S}_1, \hat{S}_2] = -c \tag{45}$$

is valid. Using these results, we obtain

$$E[\hat{Y}_A] = \hat{y}_A, \qquad E[\hat{Y}_B] = \hat{y}_B, \tag{46}$$

$$Var[\hat{Y}_A] = u(\hat{y}_A), \qquad Var[\hat{Y}_B] = u(\hat{y}_B), \qquad Cov[\hat{Y}_A, \hat{Y}_B] = u(\hat{y}_A, \hat{y}_B), \tag{47}$$

i. e. the estimators $\hat{Y}_A$ and $\hat{Y}_B$ yield the correct results as given by equations (27) to (31).

We observe that

$$\text{Cov}\left[\hat{S}_1, X_{A,k}\right] = \text{Cov}\left[\hat{S}_2, X_{B,k}\right] = 1 \tag{48}$$

and

$$\text{Cov}\left[\hat{S}_1, X_{B,k}\right] = \text{Cov}\left[\hat{S}_2, X_{A,k}\right] = 0. \tag{49}$$

Thus

$$u(\hat{y}_A, x_{A,k}) = \text{Cov}\left[\hat{Y}_A, X_{A,k}\right] = u^2(\hat{y}_A) \tag{50}$$

and

$$u(\hat{y}_B, x_{B,k}) = \text{Cov}\left[\hat{Y}_B, X_{B,k}\right] = u^2(\hat{y}_B). \tag{51}$$

This is what we wanted to show.

Appendix 2 (Uncorrelated measurement results)

Here we consider the possibility that the measurement results of the linking laboratories (group C) have to be treated as if they were uncorrelated, because the respective covariances have not been reported. In this case we have to set all covariances $u(x_{A,i}, x_{B,i})$ $(i \in (\mathfrak{I}_A \cap \mathfrak{I}_B))$ equal to zero in the equations (22) to (26). This yields

$$a = \sum_{i \in \mathfrak{I}_A} \frac{1}{u^2(x_{A,i})}, \tag{52}$$

$$b = \sum_{i \in \mathfrak{I}_B} \frac{1}{u^2(x_{B,i})}, \tag{53}$$

$$c = 0, \tag{54}$$

$$s_1 = \sum_{i \in \mathfrak{I}_A} \frac{x_{A,i}}{u^2(x_{A,i})}, \tag{55}$$

$$s_2 = \sum_{i \in \mathfrak{I}_B} \frac{x_{B,i}}{u^2(x_{B,i})}. \tag{56}$$

Thus, we obtain from equations (27) and (28) for the key comparison reference values

$$\hat{y}_A = \frac{s_1}{a} \qquad \text{and} \qquad \hat{y}_B = \frac{s_2}{b}, \tag{57}$$

and from equations (29) and (30) for their associated standard uncertainties

$$u(\hat{y}_A) = \frac{1}{\sqrt{a}} \qquad \text{and} \qquad u(\hat{y}_B) = \frac{1}{\sqrt{b}}, \tag{58}$$

i. e. the usual weighted mean results of for the KCRVs separately for each group participating in the key comparisons. The results of the groups are independent of each other as indicated by the covariance of the KCRVs, $u(\hat{y}_A, \hat{y}_B) = 0$, as obtained from equation (31). This demonstrates that the linking procedure essentially depends on the covariances reported by the linking laboratories.

References

[1] Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, International Committee of Weights and Measures (CIPM), 14 October 1999, (http://www1.bipm.org/utils/en/pdf/mra_2003.pdf).

[2] Guidelines for CIPM key comparisons, International Committee of Weights and Measures (CIPM), 1 March 1999, (http://www1.bipm.org/utils/en/pdf/guidelines.pdf).

[3] KCDB, http://www.bipm.org/en/cipm-mra/mra-kcdb/

[4] Steele A G, Wood B M and Douglas R J 2005, *Linking key comparison data to appendix B*, TEMPMEKO 2004 : 9th International Symposium on Temperature and Thermal Measurements in Industry and Science (Dubrovnik, Croatia, June 22-25, 2004), p. 1087-1092, vol. 2

[5] Nielsen L 2000, *Evaluation of measurement intercomparisons by the method of least squares*, Danish Institute of Fundamental Metrology, Technical Report DFM-99-R39

[6] Elster C, Link A and Wöger W 2003, *Proposal for linking the results of CIPM and RMO key comparisons*, Metrologia **40**, 189–94

[7] White D R 2004, *On the analysis of measurement comparisons*, Metrologia **41**, 122–31

[8] Sutton C M 2004, *Analysis and linking of international measurement comparisons*, Metrologia **41**, 272–7

[9] Thalmann R 2002, *CCL key comparison: calibration of gauge blocks by interferometry*, Metrologia **39**, 165–77

[10] CCL/WGDM/09-22: http://www.bipm.org/wg/CCL/CCL-WG/Allowed/General_CCL-WG_docs/CCL-WG-MRA-GD-2.pdf, accessed 2014-07-22

[11] Decker J et. al. 2008, *Measurement science and the linking of CIPM and regional key comparisons*, Metrologia **45**, 223–232

[12] Decker J E, Altschuler J, Beladie H, Malinovsky I, Prieto E, Stoup J, Titov A, Viliesid M and Pekelsky J R 2007, *Report on SIM.L-K1 regional comparison: stage one calibration of gauge blocks by optical interferometry*, Metrologia **44** (Tech. Suppl.) 04001

[13] ISO 1993, *Guide to the Expression of Uncertainty in Measurement* (Geneva: ISO)

[14] Weise K and Wöger W 1992, *A Bayesian theory of measurement uncertainty*, Meas. Sci. Technol. **3**, 1–11

[15] Jeffreys H 1939, *Theory of Probability* (Oxford: Oxford University Press)

[16] Jaynes E T 1957, *Information theory and statistical mechanics*, Phys. Rev. **106**, 620–630

[17] Jaynes E T 1957, *Information theory and statistical mechanics II*, Phys. Rev. **108**, 171–190

[18] Jaynes E T 1968, *Prior probabilities*, IEEE Trans. On Systems Science and Cybernetics **SC-4**(3), 227–241

[19] Shannon C E 1948, *A mathematical theory of communication*, Bell System Techn. Journ. **27**, 379–423 and 623–656

[20] Kullback S and Leibler R A 1951, *On information and sufficiency*, Ann. Math. Stat. **22**, 79–86

[21] Acko B 2012, *Final report on EUROMET Key Comparison EUROMET.L-K7: Calibration of line scales*, Metrologia **49**, 04006